MULTIMODE DELIVERY OF MULTIMEDIA INFORMATION

Frank L. Walker and George R. Thoma
National Library of Medicine

Keywords: HotMed, MedJava, Internet, Java, Document Delivery, Email, MIME, Ariel, DocView, World Wide Web, FTP, Interlibrary Loan, PDF, NLM

Abstract: Increasing numbers of libraries are using the Internet for document delivery, not only for interlibrary loan, but for delivering documents directly to the patron's desktop computer. The widespread adoption of software such as Ariel and DocView is making this possible. Ariel, a product of the Research Libraries Group, converts paper-based documents to monochrome (black and white) bitmapped images, and delivers them over the Internet. The National Library of Medicine's DocView is primarily designed for library patrons to receive, display and manage documents received from Ariel systems. While libraries and their patrons are beginning to reap the benefits of this new technology, there still are areas where improvements can be made. The Communications Engineering Branch of the Lister Hill National Center for Biomedical Communications, an R&D division of NLM, is developing software code-named HotMed to increase the options for delivering library information over the Internet. For instance, HotMed adds a third method for one-time information delivery, via the World Wide Web (WWW), to the two methods already used by Ariel and DocView: File Transfer Protocol (FTP) and Multipurpose Internet Mail Extensions (MIME) Email. HotMed also permits delivery of multiple files to multiple recipients, and goes beyond monochrome bitmapped images to deliver any multimedia file type, such as color and grayscale images, audio and video information, and word processing documents. Finally, by incorporating a Java applet (MedJava) for delivering documents via the WWW, the HotMed software gives the librarian a new option that potentially solves several problems commonly associated with Internet document delivery. This paper describes the design of the prototype HotMed software and shows how it can aid the librarian in delivering many types of library information over the Internet to the patron's desktop.

1. BACKGROUND: DOCVIEW FOR THE END USER

Document delivery by libraries and information service providers has evolved over the past twenty years. Back in the late 1970's, libraries doing interlibrary loan were beginning to adopt facsimile transmission as an electronic supplement to traditional postal mail. While delivery of library documents via mail and fax still play a major role in document delivery, the 1990's decade has seen the introduction of Internet document delivery. Several thousand libraries are using the Ariel system software distributed by Research Libraries Group for interlibrary loan via the Internet.[1,2] Ariel uses a technology that is faster than mail, more reliable than fax, and that offers higher resolution images than possible through conventional fax. The second half of this decade has also seen a growing trend for Internet delivery of library documents directly to the desktop

computers of library patrons. One software program that has helped make this possible is the National Library of Medicine's DocView.[3,4]

DocView, freely available from NLM, is software that enables a library patron to receive document images sent by a library's Ariel system. DocView's compatibility with Ariel enables a library or document supplier to use Ariel to scan a printed document and send the resulting images directly to a patron's computer running DocView under any Windows™ operating system. Ariel sends the file of bitmapped images via File Transfer Protocol (FTP)[5] protocol or Multipurpose Mime Email Extensions (MIME) email.[6] DocView is capable of displaying monochrome bitmapped images in either the Group on Electronic Document Interchange[7] (GEDI) file format used by Ariel systems, or in the Tagged Image File Format[8] (TIFF). DocView permits the user to zoom, scroll, pan and rotate document images. A user may "bookmark" pages for easy browsing, or copy the received images and paste them in word processing documents. DocView also enables the user to file and organize the received documents through a built-in document management system. Finally, DocView permits the user to forward documents over the Internet to others, using either FTP or MIME email.

A comprehensive beta test of DocView lasting 2½ years revealed that a large majority of users felt that DocView had improved the delivery of documents from their libraries.[9] As a result, DocView was released in January 1998 and is freely available. Since its release the DocView software has been downloaded by more than 2,000 registered users in over 80 countries. A web site established to distribute DocView includes an extensive user manual, a report on the DocView beta test, and published papers related to DocView. The software can be downloaded from the DocView home page on this web site: http://archive.nlm.nih.gov/proj/docview/project.htm.

2. NEW DELIVERY OPTIONS

While the Internet and products such as Ariel and DocView have changed the nature of library document delivery, improvement is possible in several areas. First, the Ariel system allows only delivery of monochrome images (e.g., printed black and white pages), and not grayscale or color images. Second, it does not permit delivery of alternative multimedia file formats, such as video or audio information, or word processing documents. Third, it restricts document delivery to FTP and MIME email transmission, each of which has shortcomings. Fourth, Ariel running on a single computer is not designed to scale up for heavy delivery loads.

In line with our mandate to continue R&D in document delivery, we are developing software code-named HotMed that will increase the options for document delivery available to libraries and document suppliers. HotMed will permit one-time delivery over the Internet of a wide variety of files. It will incorporate delivery through the World Wide Web, in addition to FTP and MIME email delivery, to give three choices in Internet delivery mode. It will also scale up easily so that multiple computers may be combined to handle heavy delivery loads at large libraries. By providing multimode delivery of

multimedia information, the HotMed software opens up new opportunities for delivery of library information. Table 1 compares the three delivery modes.

Functions / Features	Ariel FTP	MIME Email	WWW
Delivery to desktop possible without patron intervention	Yes	Yes	No
Intermediate server required for storage	No	Yes	Yes
Receiving computer needs to run for long periods of time	Yes	No	No
Firewalls can prevent flow of information	Yes	No	Yes
Delivery of multimedia information	No	Yes	Yes
Delivery of multiple files to multiple patrons	No	Yes	Yes
Requires a fixed IP address	Yes	No	No
Document viewer required on receiving computer	Yes	Yes	No

Table 1. Comparison of HotMed Delivery Methods

A detailed discussion of the items in Table 1 follows:

- Delivery to desktop without patron intervention

FTP and MIME email delivery can be accomplished without patrons initiating the transfer. This allows the delivered information to be immediately available for use. For FTP delivery, DocView at the end user's machine runs as a background task and automatically notifies the patron when a new document is available. Similarly, many email client software programs can automatically call an email server periodically and download messages in the background. Delivery through the WWW requires that the patron use a browser to visit a web site and manually initiate the file download. Depending on the size of the file and the communication speed, information downloaded from the web may not be available for use for several minutes.

- Intermediate server required for storage

An advantage that FTP delivery has over the two alternatives is that it does not require an intermediate server. Because of this it is the most direct delivery method, in which documents are sent directly from the Ariel computer to the DocView computer. While email delivery requires an intermediate email server, WWW delivery requires a web server. Many organizations place restrictions on email delivery, impacting the ability to deliver library documents via email. For example, some email services limit the size of email attachments to 500 kilobytes. This would preclude the delivery of a typical ten page scanned journal article, which averages one megabyte in size. A second restriction placed by email servers is that disk space allocated to user accounts is often limited in size. This limits the number of document image files that may be stored in a user's account on the email server. On the other hand, web delivery via HotMed is designed to overcome the limitations posed by email servers. The only build-up of files occurs on the web server. However, to mitigate this potential problem, HotMed can be set to

automatically remove files a nominal two days after the patron downloads them from the web server.

- Receiving computer needs to run for long periods of time

A chief drawback of FTP document delivery is that the Ariel system and the DocView computer need to run simultaneously. This means that the computer running DocView should be on as much as possible. In some cases this may prove to be an inconvenience for the patron. Email and web delivery are somewhat different, because in these situations the patron can govern the time at which delivery takes place. In these cases the receiving computer needs to be running only during the user-determined time for information delivery.

- Firewalls can prevent flow of information

Firewalls can prevent flow of information sent via FTP. Some firewalls can also restrict some forms of web delivery, especially the delivery of Java applets such as that used by HotMed. In this case, the librarian using HotMed can choose an alternative file format (TIFF or PDF) or delivery mode that would be able to cross the firewall. Email delivery has a distinct advantage here over FTP, since firewalls normally permit email to flow into and out of an organization.

- Delivery of multimedia information

The Ariel system restricts both FTP and MIME email delivery to monochrome bitmapped images. However, there is nothing inherent in the FTP or email protocols that restrict file type. Because HotMed is designed to be compatible with Ariel for FTP delivery, it also restricts the file type for this delivery mode. On the other hand, HotMed places no such restriction on email or WWW delivery. Any multimedia file may be delivered via HotMed's email facility or via the WWW.

- Delivery of multiple files to multiple patrons

Compatibility with the Ariel system for FTP delivery restricts HotMed to delivery of just one file at a time to one recipient. While Ariel also places this restriction on MIME email delivery, HotMed does not do so. HotMed can deliver multiple files to multiple recipients via both MIME email and WWW.

- Requires a fixed IP address

A major drawback of FTP delivery is that the receiving computer must have a fixed Internet Protocol (IP) address. This prevents delivery to most patrons who have dialup Internet connections, since in this case IP addresses are usually randomly assigned to the patron's computer when it establishes an Internet connection. It is not possible for the document delivery librarian to know the IP address in advance of delivery for most dial-up connections. Email and web deliveries do not have this problem, since the IP address is not a factor in delivery.

- Document viewer required on receiving computer

The web's chief advantage over the two delivery alternatives is that a document viewer is not necessarily required for WWW document delivery. For delivery of document images, HotMed's design permits it to deliver a Java applet code-named MedJava[10] with a document to the patron's web browser. The MedJava applet provides the image rendering usually supplied by a document viewer. MedJava is delivered automatically to the patron's computer, and unlike conventional document viewers, does not require user installation.

3. HOTMED IMPLEMENTATION

The implementation for HotMed running on a single computer is illustrated in Figure 1. The computer may be running under any of the Windows 95, 98, NT Workstation or NT Server operating systems.

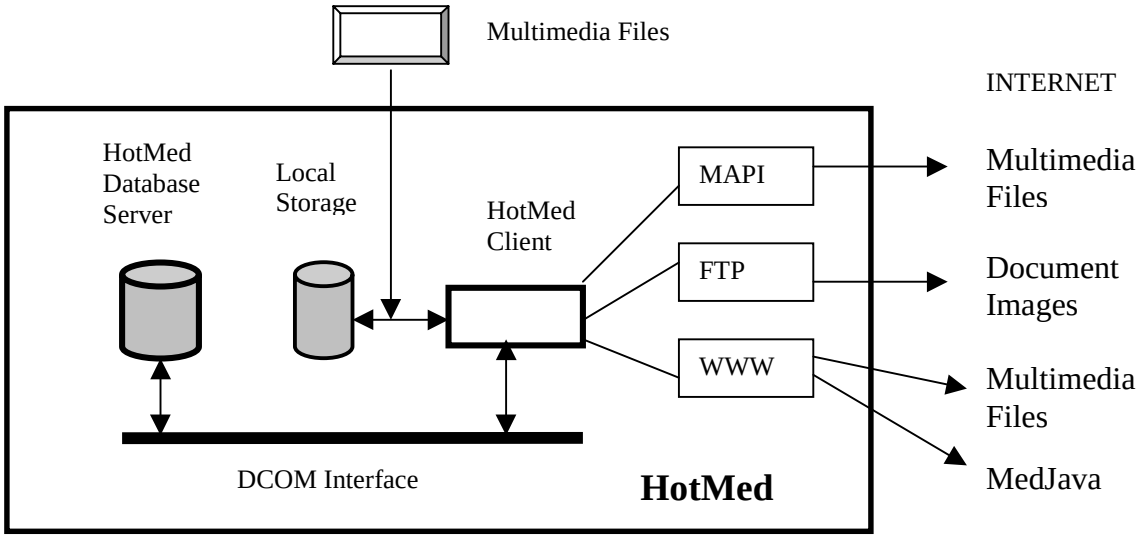

Figure 1. HotMed Configuration

This is a brief description of the components in HotMed:

- Multimedia Files: Files to be delivered to the patron. Included are video, audio, document images both monochrome and color, and word processing documents.
- HotMed Client: Controls all delivery and provides the user interface.
- HotMed Database Server: A database that keeps track of all deliveries.
- DCOM Interface: A Windows software mechanism called Distributed Component Object Model that allows communication among software components.[11]
- Local Storage: A hard disk drive on the computer for temporarily storing document images for delivery.
- MAPI: Windows messaging system called Messaging Application Protocol Interface[12] that provides MIME email delivery to patrons. For Web delivery, MAPI provides the method for notifying the patron that the delivery is available on the Web server.
- FTP: Software module for sending documents via the File Transfer Protocol
- WWW: Web server running on the same computer as HotMed. Choices include the Personal Web Server (Windows 95 and 98), Peer Web Services (Windows NT) or Internet Information Server (Windows NT Server).
- MedJava: NLM's Java applet delivered via the WWW with scanned TIFF image documents.

The HotMed delivery system can be easily expanded, as information delivery needs rise. A large number of deliveries via FTP or the web may necessitate adding more computers

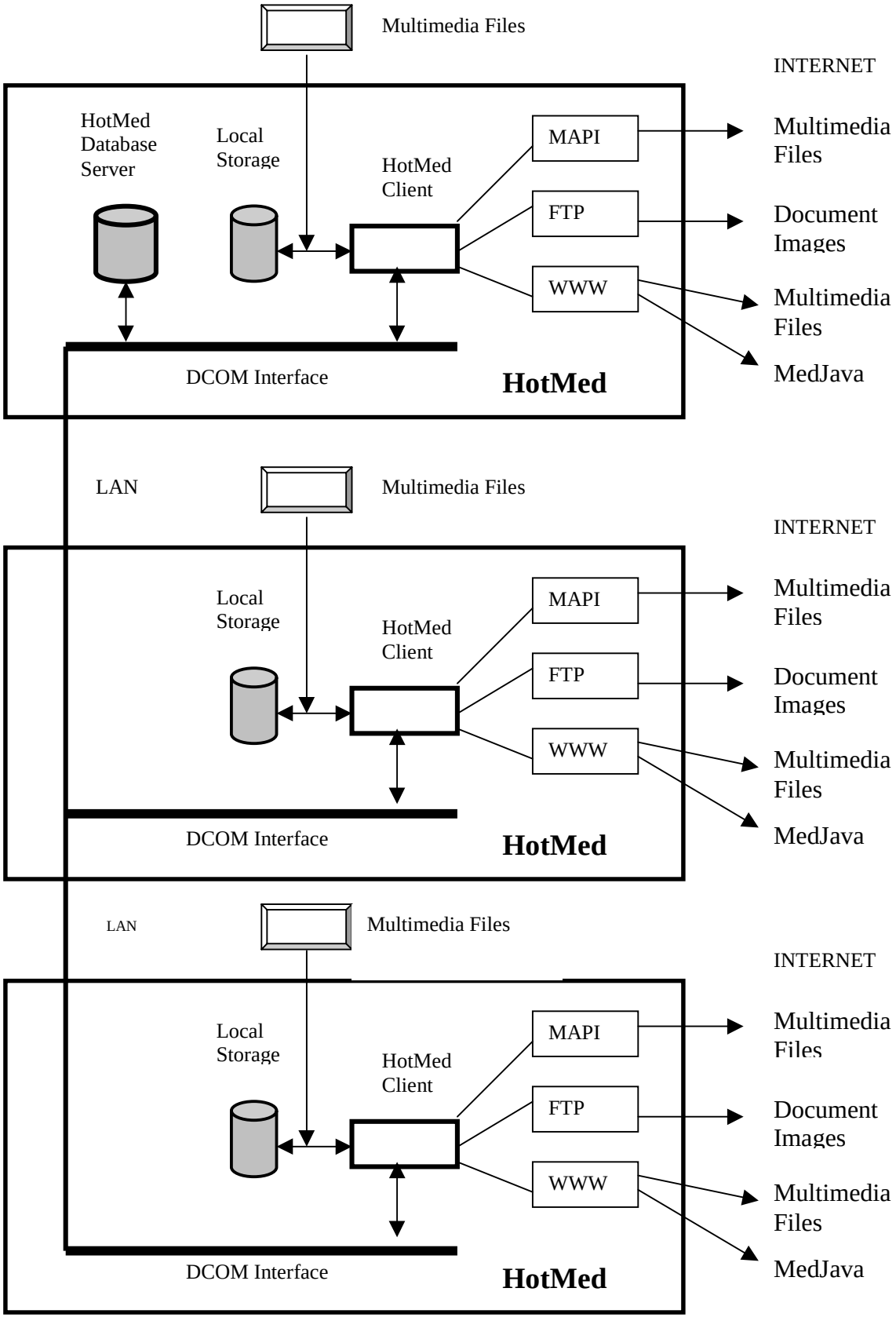

Figure 2. Multiple HotMed Computer Configuration

to increase throughput and alleviate the load. Figure 2 illustrates how three computers may be connected together in a local area network. In this configuration, any of the computers serves as a deliverer of information. Each computer can be configured to deliver the information in any of the three modes. For example, one computer could serve FTP delivery, one email delivery, and all three could serve WWW delivery. A shared database (HotMed Database Server in Fig. 2) that is maintained on one of the computers keeps track of deliveries for the entire system.

Information delivery takes place in these steps:

1. The document delivery librarian using HotMed chooses the files to be delivered to the patron. The files may be imported from any source. For example, if FTP delivery is desired, they may be scanned images from an Ariel system. For MIME email or Web delivery, the files may include any multimedia format, such as word processing documents, audio or video information.
2. Using a built-in address book, the operator selects the recipient's email address (for MIME email or WWW delivery) or IP address (for FTP delivery).
3. The operator decides how the information is to be delivered: Ariel FTP, MIME email or World Wide Web. Once the method of delivery is chosen, the HotMed client places routing information in the HotMed Database Server. The Database Server keeps track of which file is being delivered to which recipient.
4. The HotMed Client delivers the file(s) via the specified mode. For FTP delivery, the document is transmitted to an Ariel system or DocView running on the recipient's computer. Files delivered via MIME email are sent as attachments to the MIME email message. The email remains at the patron's email server until the patron logs in and retrieves the email. If a file is delivered via the WWW, the HotMed client software first places the file on the web server running on the HotMed computer. It then notifies the recipient through email of the delivery, and it provides the patron with the web location (URL) where the file is located.
5. For web delivery, when a fixed amount of time elapses (nominally 48 hours) after the patron first downloads the document from the Web server, HotMed removes the file from the local storage of the HotMed computer so as not to overload the system with older files. At the option of the operator, the files may be maintained in local storage for a nominal 21 days if the patron has not retrieved them, and then HotMed will delete them at that point.

In addition to the TIFF (or GEDI) file format delivered by Ariel systems, HotMed can deliver scanned document images using two alternatives: Adobe's Portable Document Format™ (PDF) and MedJava. The HotMed operator selects the document format prior to delivery, and HotMed automatically converts the file from TIFF to PDF, or from TIFF to MedJava. Table 2 lists the options available for document delivery via HotMed.

	Method of Delivery		
	Ariel FTP	**MIME Email**	**WWW**
Document Type	TIFF	TIFF, PDF	TIFF, PDF, MedJava

Table 2. HotMed Delivery Methods and Document Types

The advantage of using PDF over TIFF as a format for document delivery is that viewers are freely available for it on three major computing platforms: Windows, Macintosh and UNIX. The Adobe Acrobat Reader™ works uniformly well on all three platforms. On the other hand, the availability of TIFF viewers for the three computing platforms is not consistent. DocView and other TIFF viewers are freely available for the Windows platform, but users of Macintosh and UNIX computers typically need to go to either shareware distributors or commercial vendors to get suitable viewers for multipage TIFF files. The fact that Adobe Acrobat Reader is freely available from one source for all three platforms simplifies the task of a document delivery librarian who needs to equip a patron population with document viewing software.

The primary advantage of Java-assisted document delivery over TIFF and PDF is that it promises to eliminate the requirement for a patron to acquire and install document-viewing software on the computer. The patron needs only a web browser that is Java-enabled, e.g., capable of running a Java applet within the browser. Most desktop computers sold today come equipped with Java-enabled web browsers. HotMed's web-based delivery method takes advantage of this by delivering a Java applet along with the document images. The MedJava applet renders the document images within the browser's window and provides the interface for using the document.

Running inside the patron's web browser, MedJava allows the patron to retrieve, view and print the document located on the web server. MedJava handles monochrome TIFF files, either uncompressed or compressed. Image expansion algorithms implemented by the applet include Packbits and ITU recommendations for 1-dimensional Group 3, 2-dimensional Group 3, and Group 4 compression.[13] To help minimize copyright abuse, the MedJava applet does not store the received documents images on the patron's hard disk. Nor does MedJava permit the user to copy, alter or retransmit the document. However, it does permit printing the document. Research learned through the DocView beta test revealed that nearly all users print received library documents. Because of the importance of printing, this capability was designed into MedJava. The software does not restrict what the user may do with the document once it is printed.

There are a number of potential problems posed by conventional document viewers to document delivery librarians.[10] HotMed is designed to mitigate these problems through its delivery of the MedJava applet. Specifically, it

1. Permits document delivery to multiple computing platforms (Windows, UNIX and Macintosh). Rather than requiring a document viewer to see the received documents, patrons need only have a web browser such as the ones from Microsoft or Netscape. These browsers run on all three computing platforms. Library documents are delivered along with the applet, which provides image rendering within the context

of the browser. As a complement to Java-assisted delivery, document delivery to multiple platforms is also improved through HotMed's automatic optional conversion of TIFF files to PDF. This provides an excellent alternative solution for document delivery to multiple platforms, especially in environments that restrict delivery of Java applets across firewalls.

2. Eliminates problems associated with software distribution and installation. Document viewers need to be provided to library patrons and installed on their computers, often a time-consuming process. The MedJava applet solves this problem because it installs automatically and runs on the patron's computer without user intervention.
3. Eliminates the need for the end user to update the document viewer software. As new versions of document viewers are released, they may need to be installed on the patron's computer. MedJava eliminates problems associated with version control, since the correct version will always be delivered with the library document.
4. Reduces the need for user training and documentation. Document delivery librarians may sometimes need to train users in using document viewers, and provide them written user manuals. The MedJava applet is designed to reduce or eliminate this problem, since its user interface is very simple. Users do not need to learn another software program; they need only to use their web browser, with which they are most likely already familiar.
5. Partially solves the problem of copyright and protection of intellectual material. One problem associated with all document viewers is that an electronic version of the document is sent to the recipient's computer, where it can be stored on hard disk, copied, modified or redistributed. MedJava's design addresses this issue, since it prevents the recipient from saving, copying, altering or retransmitting the received document. As already mentioned, the applet permits the user to print the received document, and it makes no attempt to prevent the user from printing multiple copies of the document.

4. HOTMED EVALUATION

After the completion of the software development, HotMed will be alpha tested in-house and beta tested externally by a number of libraries. Preliminary in-house testing has centered on performance. One aspect of interest is the relative performance of image expansion speed using a conventional document viewer versus Java-assisted viewing. Table 3 compares the image processing performance between MedJava and Windows-based DocView. To make this test, 10 sample images of a scanned biomedical journal were compressed Group 4 and stored on disk. The average time to expand and scale the images for display was calculated, first using DocView's 16-bit assembly language image processing library on a 333 MHz Windows NT workstation. Then measurements were taken on the same computer using MedJava while running under web browsers from Microsoft, Netscape and Sun Microsystems. Finally, to check the relative performance of MedJava on other platforms, it was tested on a Macintosh computer and an X-Terminal connected to a UNIX Server.

	Average Image Processing Speed Expansion + Scaling for 10 Images
DocView (16-bit assembly language)	.179 sec (333 MHz Windows NT Workstation)
MedJava (Microsoft Internet Explorer)	.428 sec (333 MHz Windows NT Workstation)
MedJava (Netscape Navigator)	.462 sec (333 MHz Windows NT Workstation)
MedJava (Sun Java Virtual Machine)	.569 sec (333 MHz Windows NT Workstation)
MedJava (Microsoft Internet Explorer)	1.670 sec (Macintosh Power PC 9500/132 MHz, OS 8.1)
MedJava (Netscape on UNIX X-Terminal)	13.360 sec (Sun Enterprise 4000 UltraSparc with 2 CPUs, each running at 266 MHz; Solaris operating System)

Table 3. Comparison of Performance of DocView versus MedJava

This table shows that on a 333 MHz PC platform it is possible to get excellent image processing performance regardless of the web browser. As expected, DocView's assembly language image processing library runs the fastest, with Java in a close second place. The results show that the Java Virtual Machine (JVM) in the Microsoft Internet Explorer is faster than that in the Netscape browser, with Sun's implementation falling in third place. These results also show that the platform of choice for running MedJava is a fast PC. Coming in second place at a slower speed is the Macintosh. The Macintosh speed is actually quite respectable, since the computer is a RISC machine running at 132MHz, nearly three times slower than the NT machine. Trailing in a distant third place is the UNIX box. The performance of UNIX running on Sun hardware is attributed to the fact that the Netscape browser did not contain a Just-In-Time (JIT) compiler; instead the bytecodes in MedJava were interpreted. The use of a JIT compiler significantly improves performance, as indicated by the Windows-based browsers.

Other issues to be addressed during in-house alpha testing will include these:

1. Is the software compatible with all 32-bit Windows platforms (e.g., Windows 95, Windows 98 and Windows NT)? To determine this, HotMed will be run on all three platforms, and all functionality thoroughly exercised. Bugs will be fixed when detected.
2. Does HotMed software scale up from one standalone computer to several networked computers without any problems? To determine this, several configurations will be tested.
3. Are there any bottlenecks created by large workloads in a multi-computer configuration? How can the design be modified to reduce or eliminate the bottlenecks? To test this, large batches of multimedia information will be delivered using combinations of each of the three delivery methods: MIME email, FTP and WWW. If it is found that one method of delivery takes an excessive amount of time, the algorithms for load distribution will be analyzed to determine whether there could be a more effective means of distributing the workload among the networked computers. If necessary, new algorithms will be incorporated to evaluate alternatives for work distribution.

Outside beta testing will be conducted and results will be obtained from a user survey that the HotMed operator will answer 45 days after HotMed is first run. The user survey will address eight areas:

1. Computer Experience and Job Function. To determine whether the ability to learn HotMed is dependent on the respondent's professional experience, and personal experience with using other Windows applications.
2. Computer Configuration. To determine the effect on performance of a single or multiple computer configuration, the operating system (Windows 95, Windows 98, Windows NT Workstation, or Windows NT Server), the speed of the computer(s), and the type and speed of the local area network.
3. Information Delivery. To find out what types of information are being delivered using HotMed, such as monochrome bitmapped images, grayscale and color imagery, audio/video information, or some other type of information.
4. Delivery Technique. To know how the operator is using HotMed to deliver information. Is it Ariel FTP, MIME email, WWW, or a combination of the three?
5. HotMed Usage. How many deliveries of each type are made on a daily and weekly basis? How likely is a user to keep using HotMed? These statistics will be derived during the beta test and will be attached to the user survey when the user answers it.
6. Impressions of HotMed. Users will be asked their subjective opinions on HotMed to find out what they think of it, and whether they desire to continue using it.
7. Learning to use HotMed. Users will be asked questions to determine whether HotMed is easy to learn to use, and whether the help facility is useful, as well as the on-line user manual.
8. HotMed Capabilities. Users will be asked specific questions on HotMed's capabilities. The purpose of this will to determine whether HotMed has functionality that is not needed, or needs to be improved. Users will be given the opportunity to suggest new functionality that they would like to see included in HotMed.

5. SUMMARY

Prototype software code-named HotMed is being developed at the R&D labs of the National Library of Medicine to give document delivery librarians new options in one-time delivery of library information. This software provides several modalities for delivery of information over the Internet to patrons. It also permits delivery of not just monochrome images produced through a scanning process, but any other type of multimedia file, including color images, word processing documents, and audio and video information. For the delivery of images from scanned documents, the HotMed software provides an option for automatic conversion of TIFF images to either PDF or MedJava format. These two formats promise to overcome some of the problems associated with delivery of TIFF files. After development this software will be evaluated through alpha and beta testing to determine its effectiveness in addressing issues related to document and information delivery through the Internet.

6. REFERENCES

1. Berger, MA. Ariel Document Delivery and the Small Academic Library. College & Undergraduate Libraries, Vol. 3(2). The Haworth Press, 1996; 49-56.
2. The World Wide Web address for Research Libraries Group is located at this URL: http://www.rlg.org.
3. Walker FL, Thoma GR. DocView: Providing Access to Printed Literature through the Internet. Proceedings IOLS'95. Medford NJ: Learned Information, 1995; 165-173.
4. Walker FL, Thoma GR. Internet Document Access and Delivery. Proceedings IOLS'96. Medford NJ: Learned Information, 1996; 107-116.
5. Postel, J. and Reynolds, J. File Transfer Protocol, Request for Comments #959, October 1985, available through the Internet at URL http://www.rfc-editor.org/.
6. Borenstein, N. and Freed, N. MIME (Multipurpose Internet Mail Extensions), Request for Comments #1341, June 1992, available through the Internet at URL http://www.rfc-editor.org/.
7. Agreements of the Group on Electronic Document Interchange (GEDI) are available through the Internet at URL http://lib.ua.ac.be/gedi/gediv21.doc
8. TIFF Revision 6.0, Aldus Corporation, June 3, 1992.
9. Walker FL, Thoma GR, "Internet Document Delivery: An End User Survey," Proc. IOLS '97. Medford N.J: Information Today, 1997; 145 - 153.
10. Walker FL, Thoma GR, "Java-Assisted Internet Document Delivery", to be published in the Proceedings of Computers in Libraries '99. Medford N.J: Information Today, 1999.
11. Sessions, R. Com and Dcom : Microsoft's Vision for Distributed Objects, John Wiley & Sons, December 1997.
12. De La Cruz, I. and Thaler, L. Inside MAPI, Microsoft Press, 1996.
13. ITU Publications are available through the Internet at this URL: http://www.itu.int/publications/bookstore.html.